Results of the AutoML challenge

Isabelle Guyon, Imad Chaabane, Hugo Jair Escalante, Sergio Escalera, Damir Jajetic, James Robert Lloyd, Nuria Macia, Bisakha Ray, Lukasz Romazco, Michele Sebag, Alexander Statnikov, Sebastien Treger, Evelyne Viegas

*automl@**ChaLearn**.org*

Thanks

Hackathon team:
Marc Boullé
Lukasz Romaszco
Sébastian Treger
Emilia Vaajoensuu
Philippe Vandermersch

Software development:
Eric Carmichael
Ivan Judson
Christophe Poulain
Percy Liang
Arthur Pesah
Xavier Baro Solé
Lukasz Romaszco
Michael Zyskowski

Advisors and beta testers:
Kristin Bennett
Marc Boullé
Cecile Germain
Cecile Capponi
Richard Caruana
Gavin Cawley
Gideon Dror
Sergio Escalera
Tin Kam Ho
Balasz Kégl
Hugo Larochelle
Víctor Ponce López
Nuria Macia
Simon Mercer
Florin Popescu
Michèle Sebag
Danny Silver,

Codalab management:
Evelyne Viegas
Percy Liang
Erick Watson

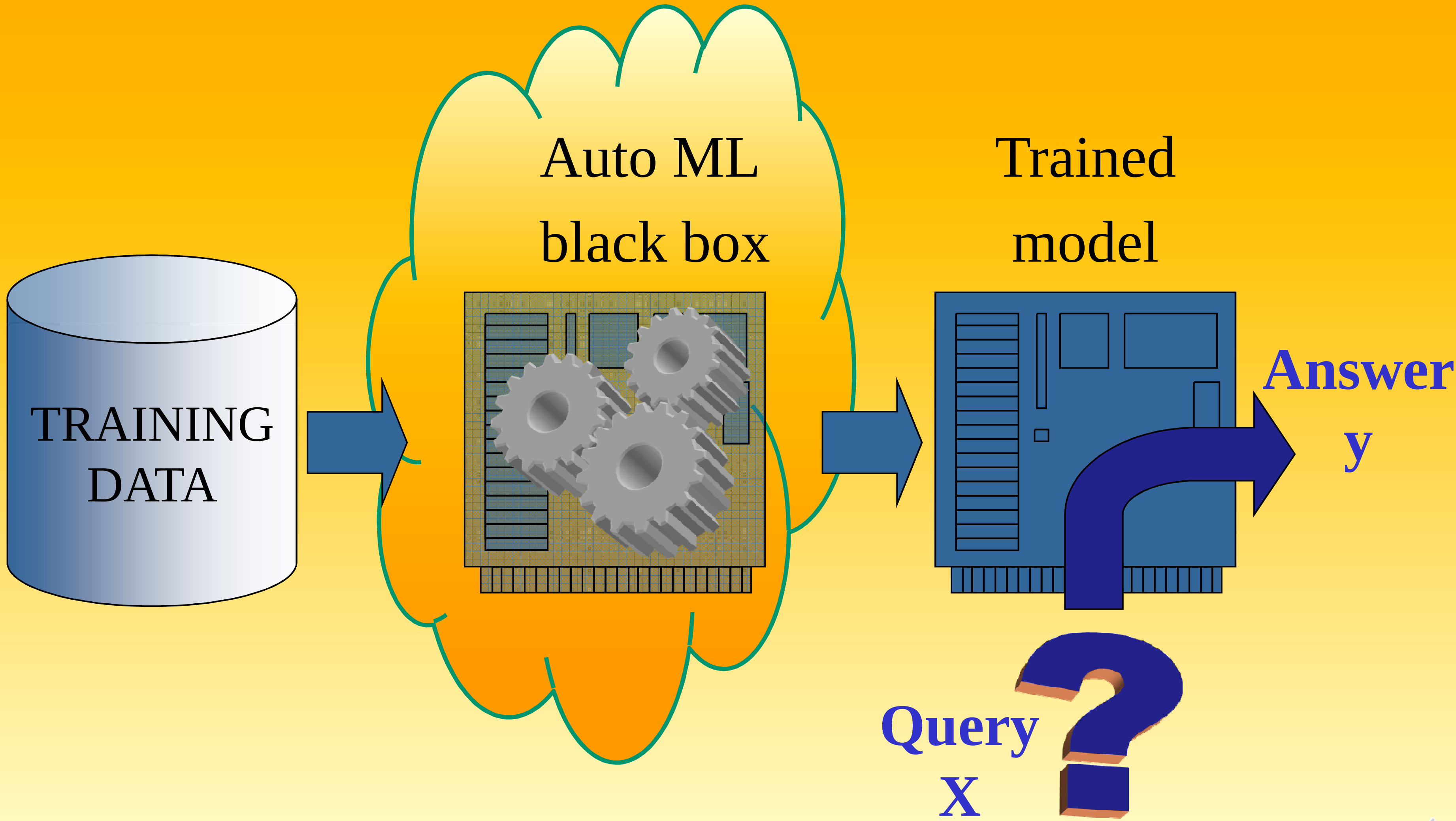
ChaLearn board ↑

Data providers:
Yindalon Aphinyanaphongs
Olivier Chapelle
Hugo Jair Escalante
Sergio Escalera
Zainab Iftikhar Malhi
Vincent Lemaire
Chih Jen Lin
Meysam Madani
Bisakha Ray
Mehreen Saeed
Alexander Statnikov
Gustavo Stolovitzky
H-J. Thiesen
Ioannis Tsamardinos

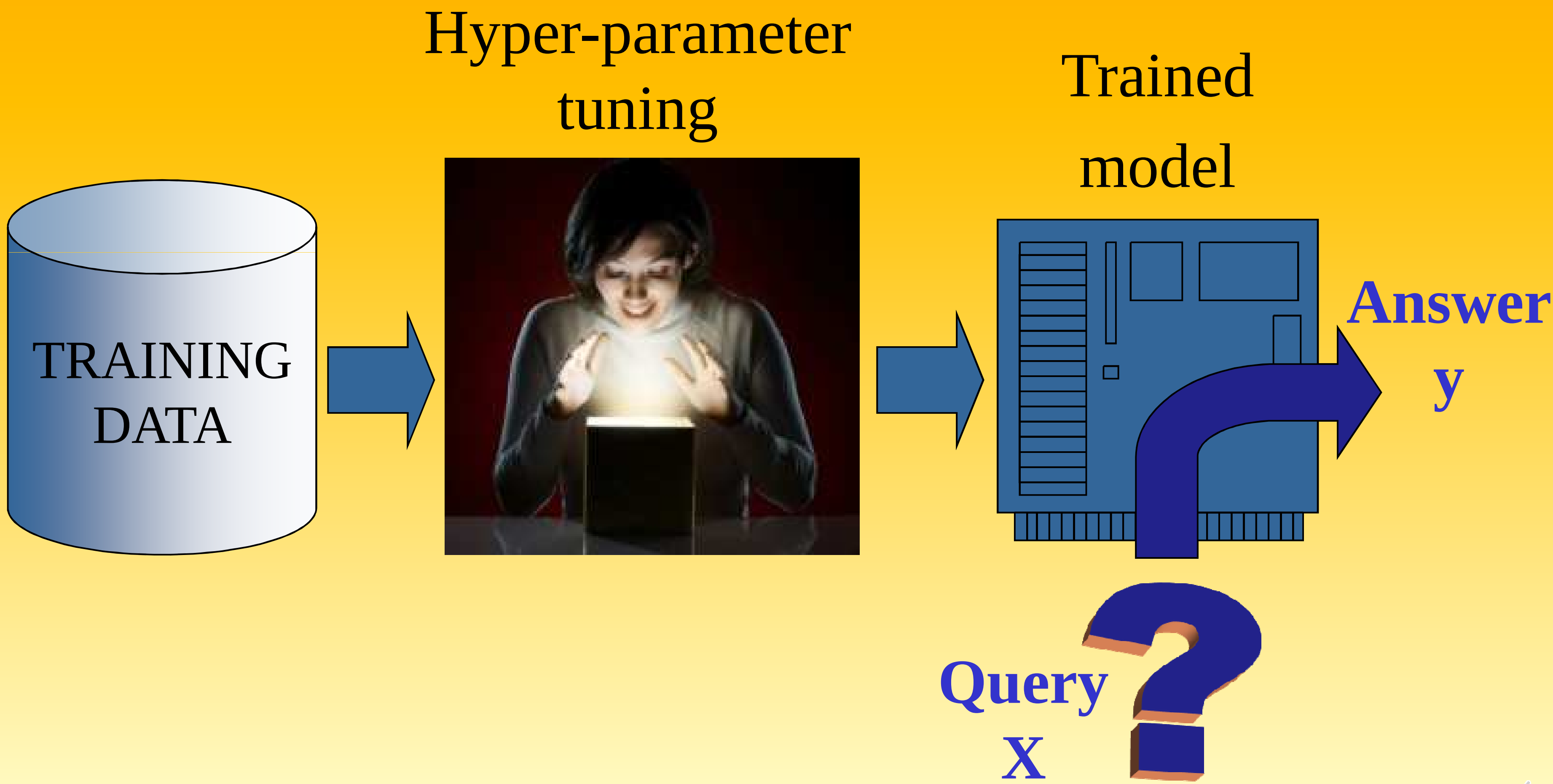
Result analysis:
Imad Chaabane

INTRODUCTION

The dream

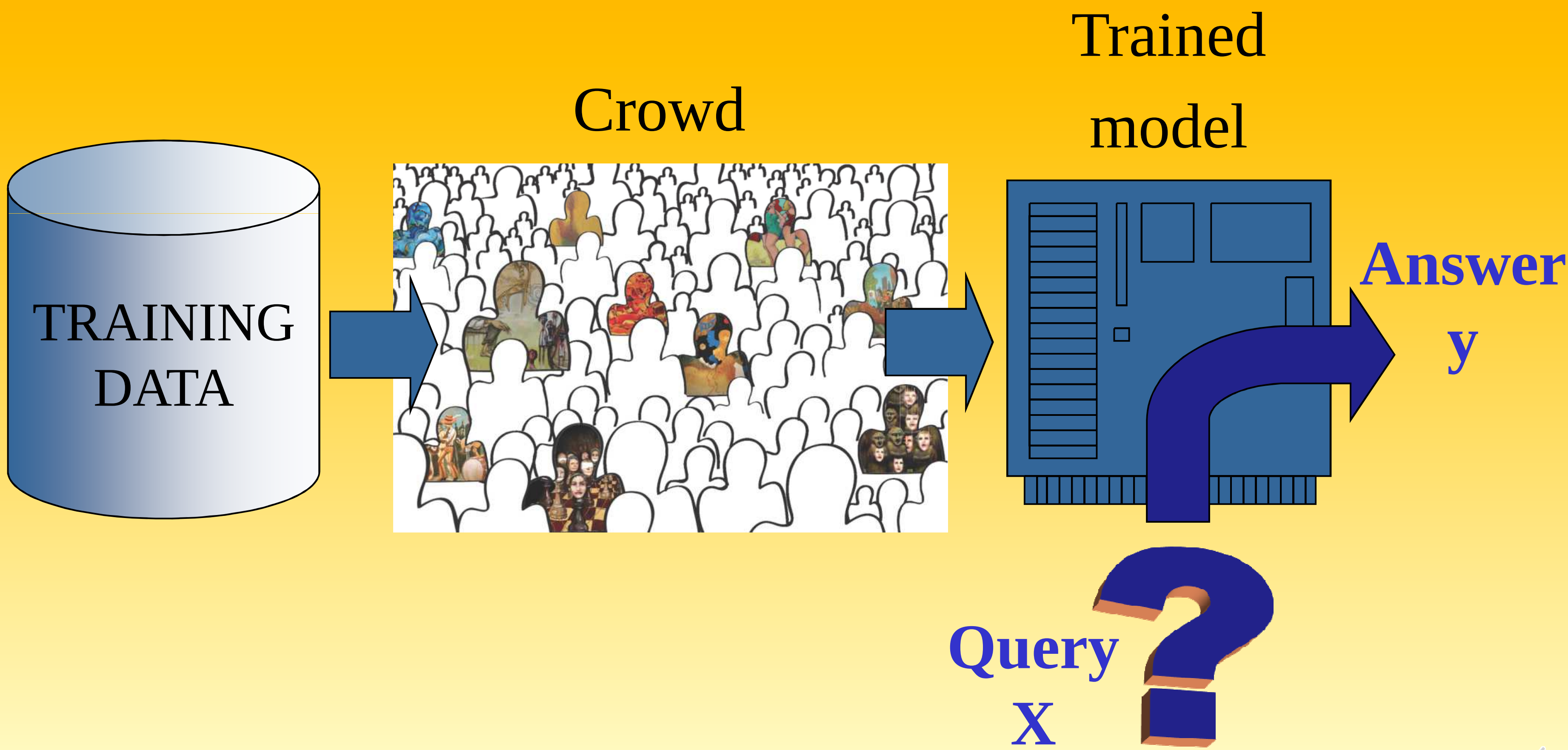

TRAINING DATA
Auto ML black box
Trained model
Answer y
Query X

The REALITY

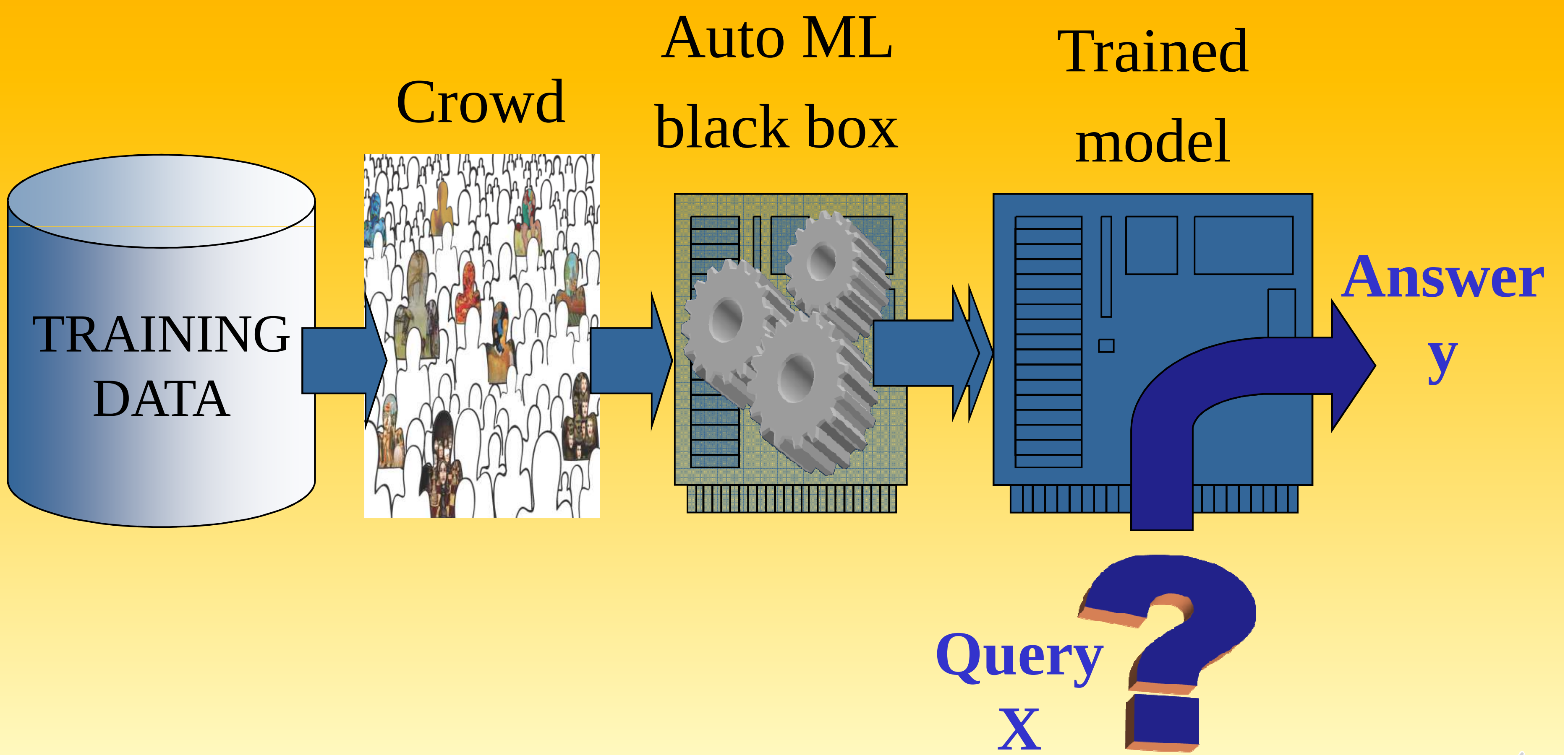


Hyper-parameter
tuning
Trained
model
TRAINING
DATA
Answer
y
Query
X

ChaLearn ML challenges



AutoML challenge



What's exciting?

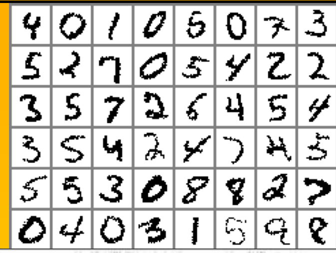
- **Intellectually challenging:**
  - Completely autonomous learner
  - Beat the "no free lunch theorem"
  - Model selection
  - Meta learning
  - Two-level objectives
  - Any overall objective (R^2, ABS, BAC, AUC, F1, PAC)
  - Any time
- **Practically important:**
  - Improve cost effectiveness
  - Improve reliability
  - Reach out to more applications

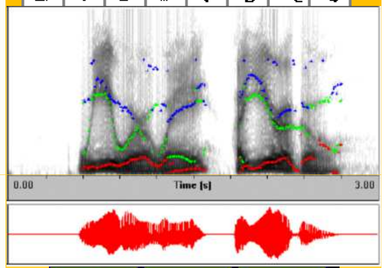
CHALLENGE DESIGN

Data: 30 large datasets

http://automl.chalearn.org/data

ound	Num	Name	Task	Metric	Time	Cnum	Cbal	Sparse	Missng	Catvar	Irrvar	Pte	Pva	Ptr	N	Ptr/N
0	1	ADULT	multilabel	F1	300	3	1	0.16	0.011	1	0.5	9768	4884	34190	24	1,424.58
0	2	CADATA	regression	R2	200	0	NaN	0	0	0	0.5	10640	5000	5000	16	312.5
0	3	DIGITS	multiclass	BAC	300	10	1	0.42	0	0	0.5	35000	20000	15000	1568	9.57
0	4	DOROTHEA	binary	AUC	100	2	0.46	0.99	0	0	0.5	800	350	800	100000	0.01
0	5	NEWSGROUPS	multiclass	PAC	300	20	1	1	0	0	0	3755	1877	13142	61188	0.21
1	1	CHRISTINE	binary	BAC	1200	2	1	0.071	0	0	0.5	2084	834	5418	1636	3.31
1	2	JASMINE	binary	BAC	1200	2	1	0.78	0	0	0.5	1756	526	2984	144	20.72
1	3	MADELINE	binary	BAC	1200	2	1	1.2 E-06	0	0	0.92	3240	1080	3140	259	12.12
1	4	PHILIPPINE	binary	BAC	1200	2	1	0.0012	0	0	0.5	4664	1166	5832	308	18.94
1	5	SYLVINE	binary	BAC	1200	2	1	0.01	0	0	0.5	10244	5124	5124	20	256.2
2	1	ALBERT	binary	F1	1200	2	1	0.049	0.14	1	0.5	51048	25526	425240	78	5,451.79
2	2	DILBERT	multiclass	PAC	1200	5	1	0	0	0	0.16	9720	4860	10000	2000	5
2	3	FABERT	multiclass	PAC	1200	7	0.96	0.99	0	0	0.5	2354	1177	8237	800	10.3
2	4	ROBERT	multiclass	BAC	1200	10	1	0.01	0	0	0	5000	2000	10000	7200	1.39
2	5	VOLKERT	multiclass	PAC	1200	10	0.89	0.34	0	0	0	7000	3500	58310	180	323.94
3	1	ALEXIS	multilabel	AUC	1200	18	0.92	0.98	0	0	0	15569	7784	54491	5000	10.9
3	2	DIONIS	multiclass	BAC	1200	355	1	0.11	0	0	0	12000	6000	416188	60	6,936.47
3	3	GRIGORIS	multilabel	AUC	1200	91	0.87	1	0	0	0	9920	6486	45400	301561	0.15
3	4	JANNIS	multiclass	BAC	1200	4	0.8	7.3 E-05	0	0	0.5	9851	4926	83733	54	1,550.61
3	5	WALLIS	multiclass	AUC	1200	11	0.91	1	0	0	0	8196	4098	10000	193731	0.05
4	1	EVITA	binary	AUC	1200	2	0.21	0.91	0	0	0.46	14000	8000	20000	3000	6.67
4	2	FLORA	regression	ABS	1200	0	NaN	0.99	0	0	0.25	2000	2000	15000	200000	0.08
4	3	HELENA	multiclass	BAC	1200	100	0.9	6 E-05	0	0	0	18628	9314	65196	27	2,414.67
4	4	TANIA	multilabel	PAC	1200	95	0.79	1	0	0	0	44635	22514	157599	47236	3.34
4	5	YOLANDA	regression	R2	1200	0	NaN	1 E-07	0	0	0.1	30000	30000	400000	100	4000
5	1	ARTURO	multiclass	F1	1200	20	1	0.82	0	0	0.5	2733	1366	9565	400	23.91
5	2	CARLO	binary	PAC	1200	2	0.097	0.0027	0	0	0.5	10000	10000	50000	1070	46.73
5	3	MARCO	multilabel	AUC	1200	180	0.76	0.99	0	0	0	20482	20482	163860	15299	10.71
5	4	PABLO	regression	ABS	1200	0	NaN	0.11	0	0	0.5	23565	23565	188524	120	1,571.03
5	5	WALDO	multiclass	BAC	1200	4	1	0.029	0	1	0.5	2430	2430	19439	270	72





Tasks

- **INPUT = I.I.D. data in feature representation, but:**
  - Sparse or full matrices.
  - Numerical/categorical/binary variables.
  - Missing values or not.
  - Noisy data or not.
  - Various proportions Ntrain / Nfeat.
- **OUPUT = one target, but:**
  - Binary (two-classes, balanced or not).
  - Categorical (multi-class: tens, hundreds of classes)
  - Multi-label.
  - Regression.
- **OBJECTIVE = miscellaneous loss funtions.**
- **COMPUTATIONAL RESOURCES = fixed (20 min on 8 core x84_64).**

Rounds

1. **NOVICE:** Binary classification.
2. **INTERMEDIATE:** Multiclass classification.
3. **ADVANCED:** Multiclass and multilabel.
4. **EXPERT:** Classification and regression.
5. **MASTER:** All of the above.

- **AutoML:** Automatic code execution on Codalab platform.
- **Tweakathon:** Result or code submission.
- To earn prizes: code should be made open source.

Protocol

ROUND **PHASE** **SUBMISSION / EVALUATION**

n-1

n

n+1

Protocol

ROUND

n-1

Blind test on fresh data **AutoML [+]** Test set

n

n+1

Protocol

ROUND **PHASE** **SUBMISSION / EVALUATION**

n-1

Code

Blind test on fresh data

AutoML [+]

Test set

n

n+1

Protocol

ROUND **PHASE** **SUBMISSION / EVALUATION**

n-1

Code

Blind test on fresh data

AutoML [+]

Data release

n

n+1

Protocol

ROUND **PHASE** **SUBMISSION / EVALUATION**

n-1

Code

Blind test on fresh data

AutoML [+]

Test set

Data release

n

Tweakathon

n+1

Protocol

ROUND | **PHASE** | **SUBMISSION / EVALUATION**

n-1

Code

Blind test on fresh data

AutoML [+]

Test set

Data release

n

Tweakathon

Code and/or **Results**

Validation set

n+1

Protocol

ROUND | **PHASE** | **SUBMISSION / EVALUATION**

n-1

Code

Blind test on fresh data

AutoML [+]

Test set

Data release

n

Tweakathon

Code and/or **Results**

Validation set

Final [+]

Test set

n+1

Protocol

ROUND	PHASE	SUBMISSION / EVALUATION	
n-1		**Code**	
Blind test on fresh data	**AutoML [+]**		Test set
Data release			
n	**Tweakathon**	**Code** and/or **Results**	Validation set
	Final [+]		Test set
Blind test on fresh data	**AutoML [+]**		***NEW*** Test set
n+1			

RESULTS

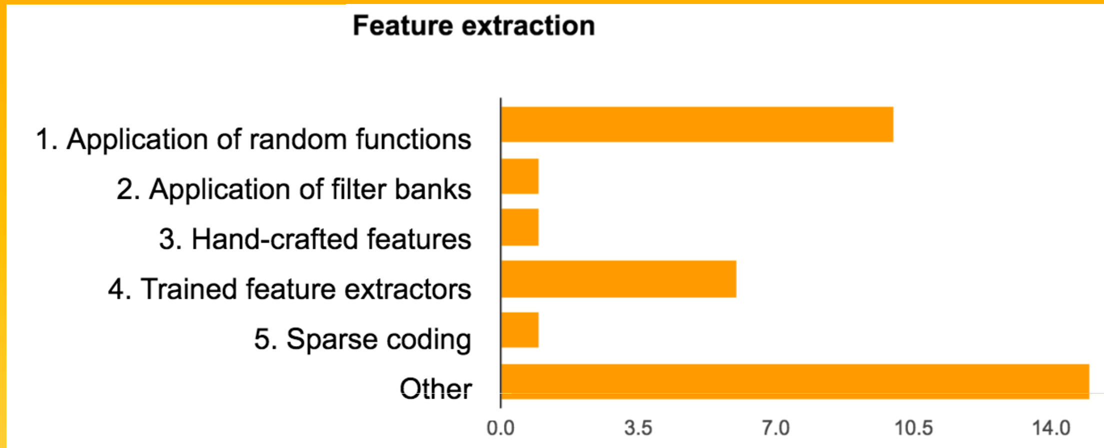
Influence of the starting kit

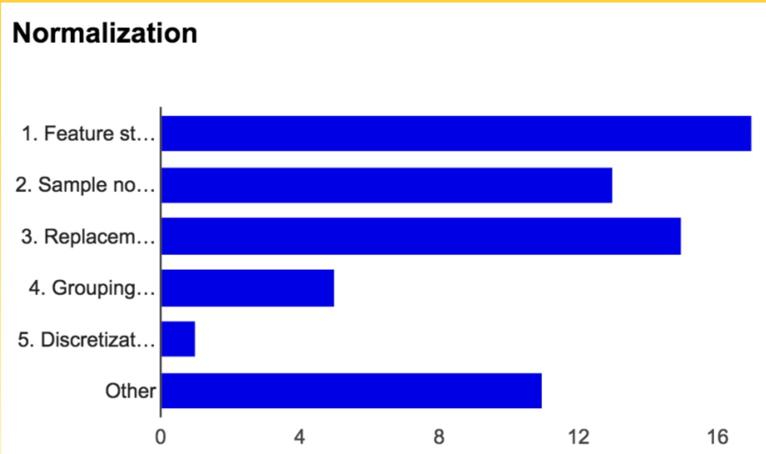
- Python example using scikit-learn: most people used it to get started.
- Many top ranking participants (including aad_freiburg) used scikit-learn.
- Codalab platform accepted:
  - Python scripts.
  - Linux executables.
  - Java JRE executables.
- C code: ideal.intel.analytics; Marc Boulle.

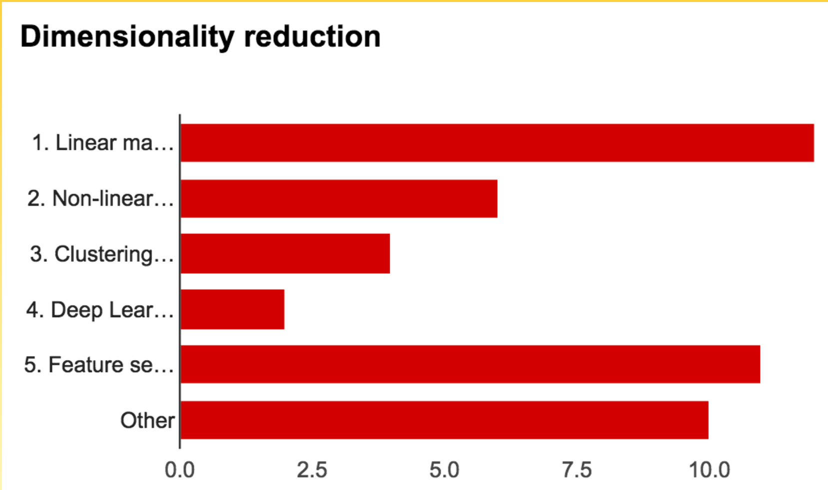
Fact Sheets (28 responses)

[link to fact sheets]

PREPROCESSING:

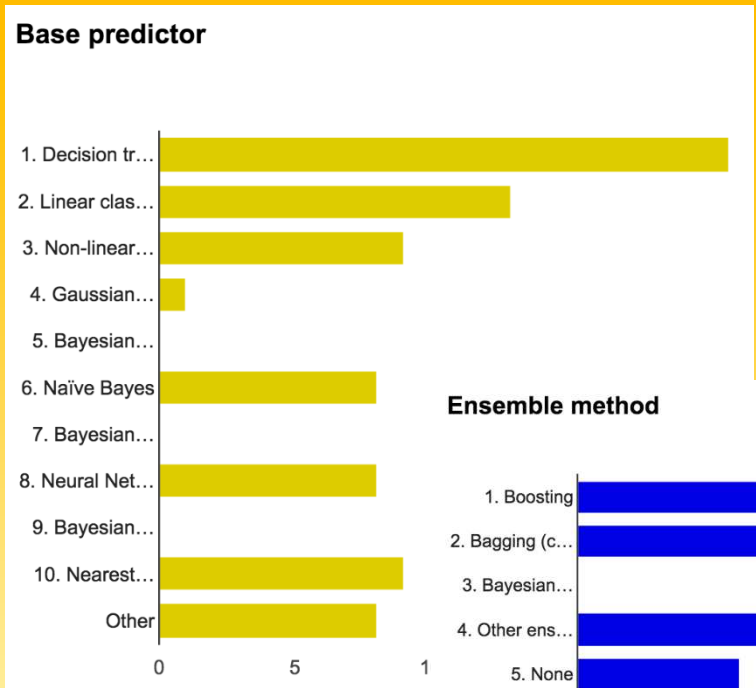


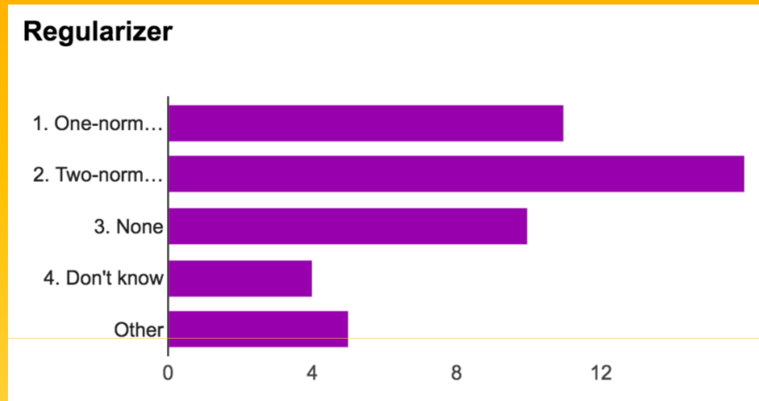


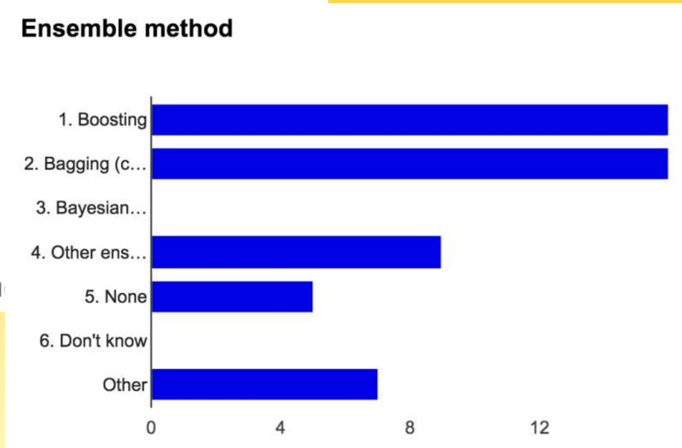


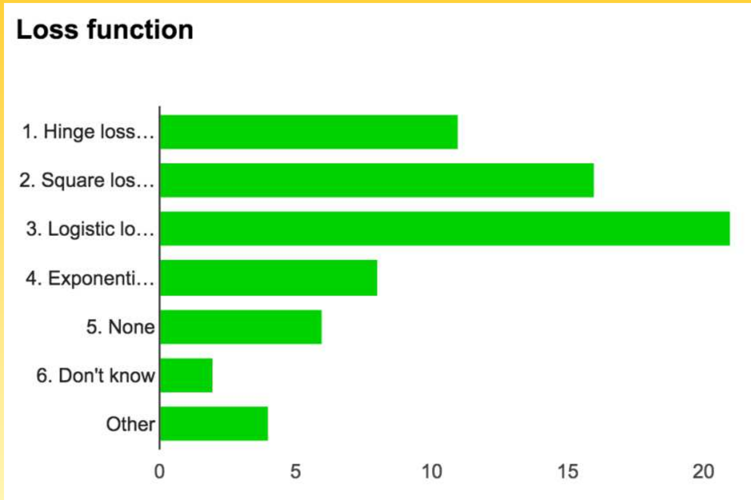
Fact Sheets (continued)

PREDICTOR:

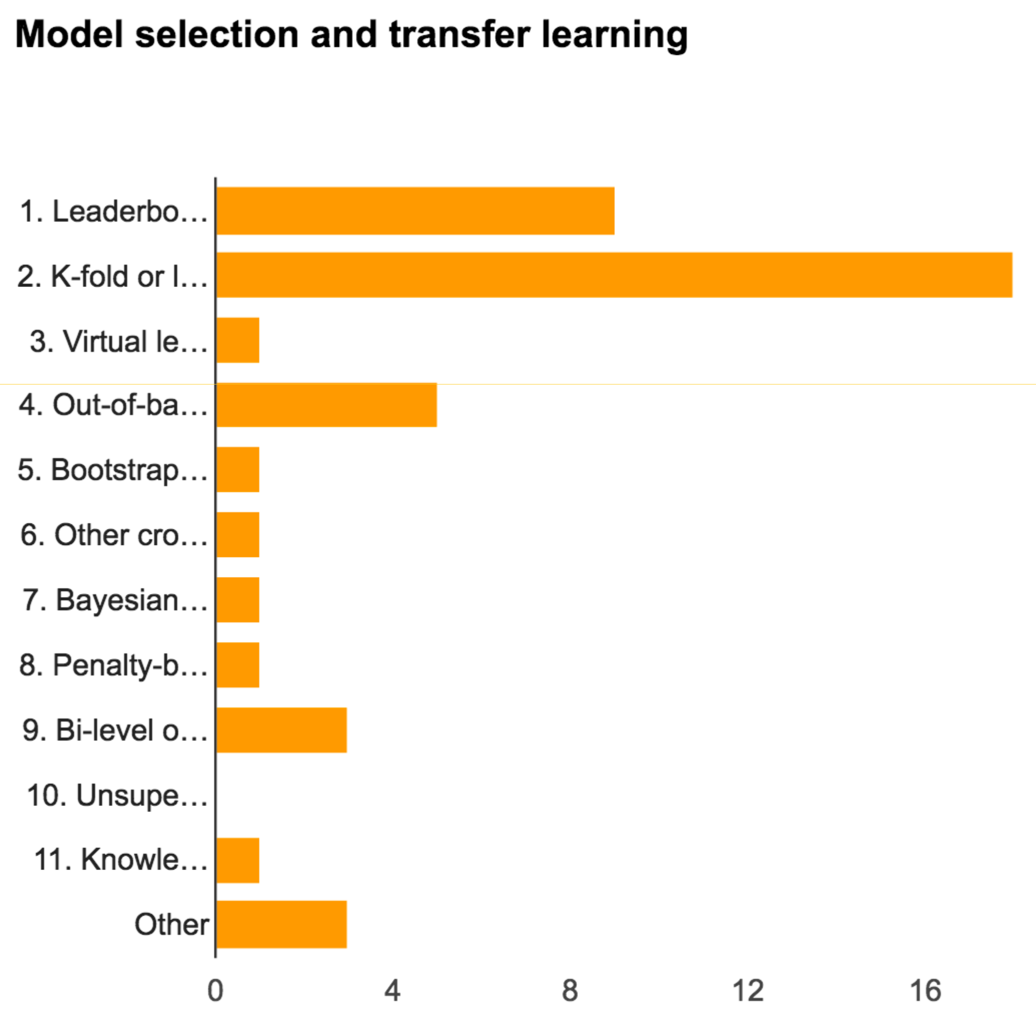




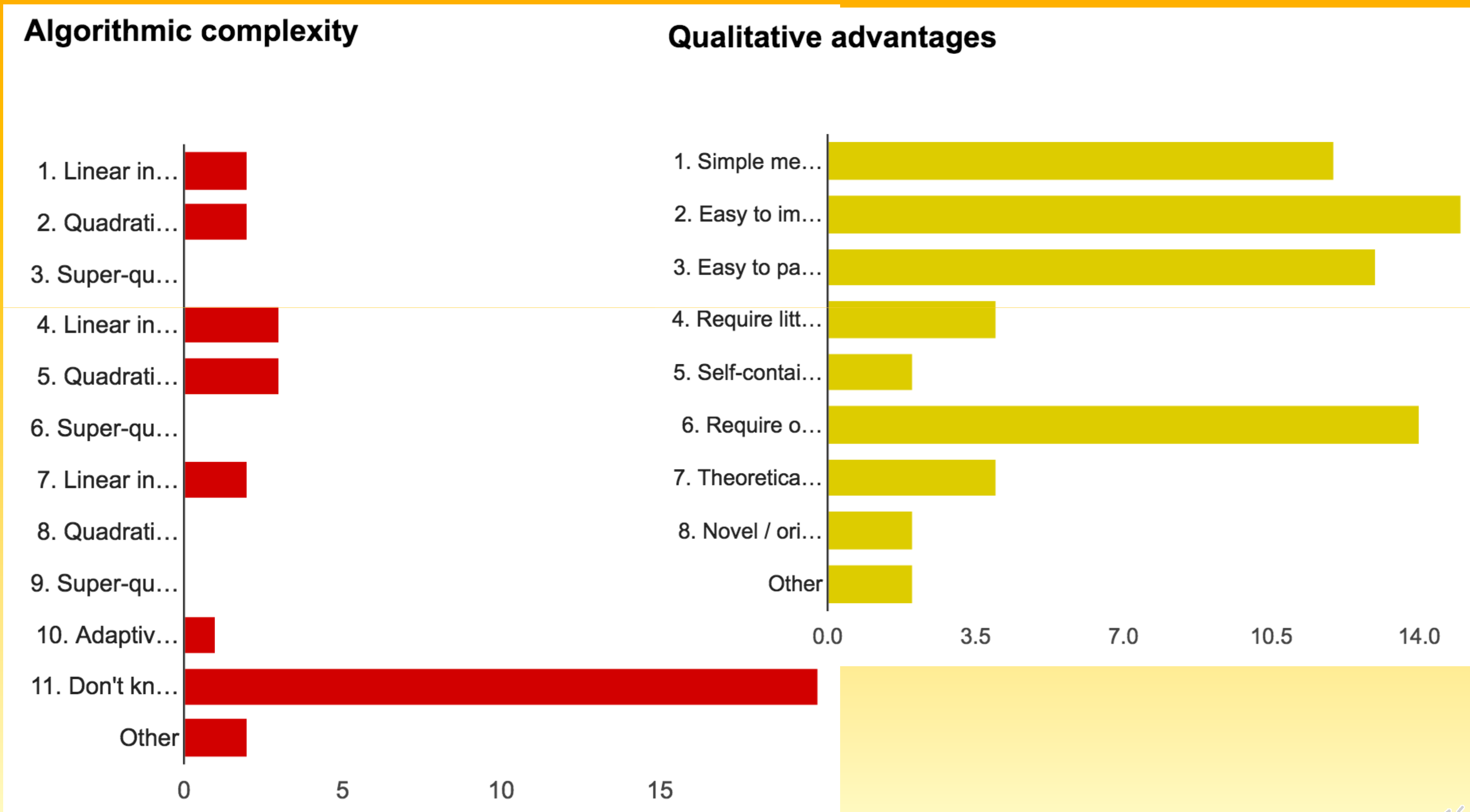




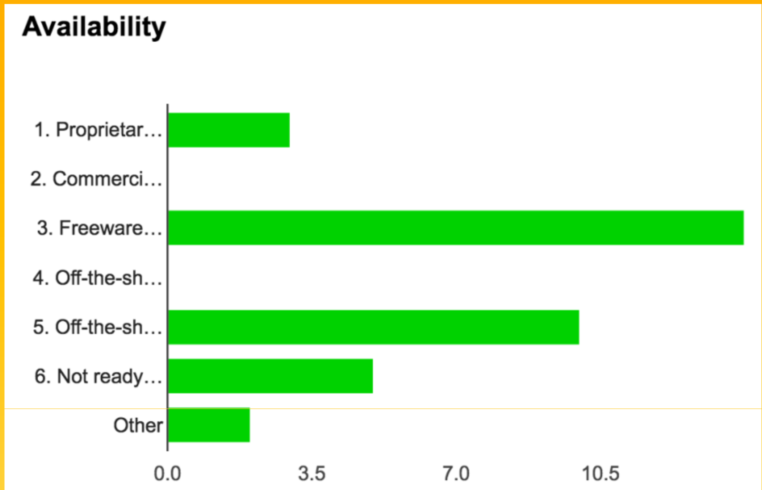
Fact Sheets (continued)

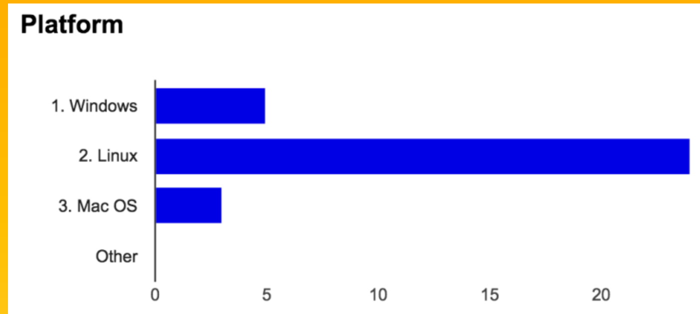


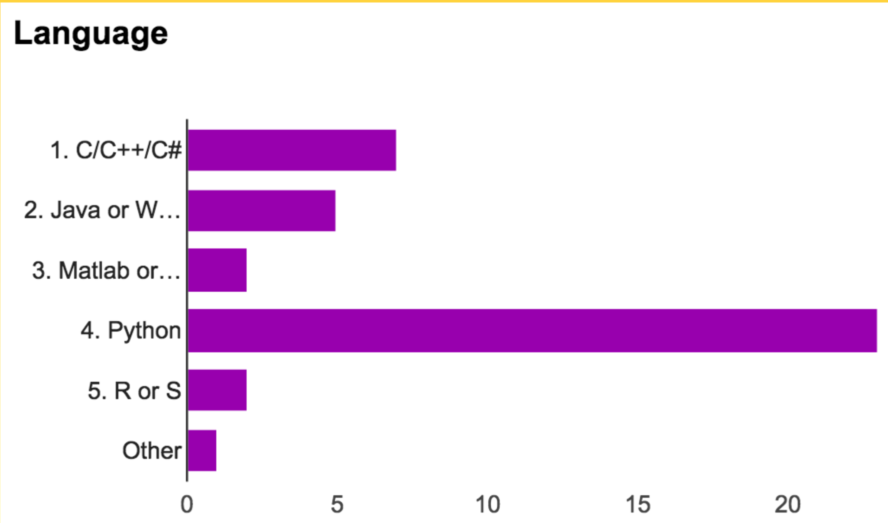
Fact Sheets (continued)



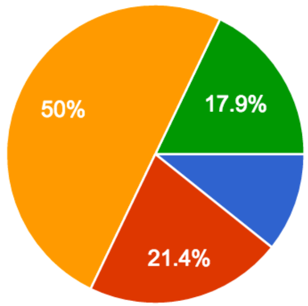
Fact Sheets (continued)







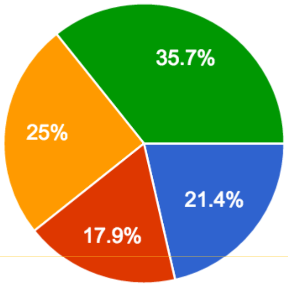
Memory



1. <= 2 GB	**3**	10.7%
2. > 2GB but <= 8 GB	**6**	21.4%
3. > 8 GB but <= 32 GB	**14**	50%
4. > 32 GB	**5**	17.9%

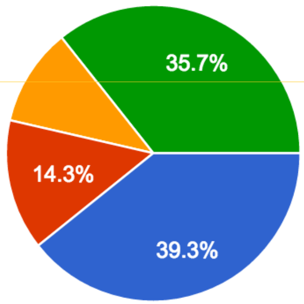
Fact Sheets (continued)

Total human effort



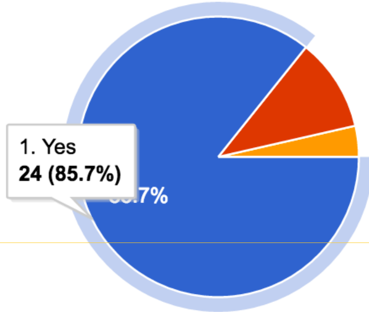
1. A few man hours	**6**	21.4%
2. A few man days	**5**	17.9%
3. 1-2 man weeks	**7**	25%
4. > 2 man weeks	**10**	35.7%

Total machine effort



1. A few hours	**11**	39.3%
2. A few days	**4**	14.3%
3. 1-2 weeks	**3**	10.7%
4. > 2 weeks	**10**	35.7%

Challenge duration OK?



1. Yes
2. No, but I cannot spend more time
3. No, please extend or run another round

Who won?

Everybody

Best overall AutoML: aad_freiburg

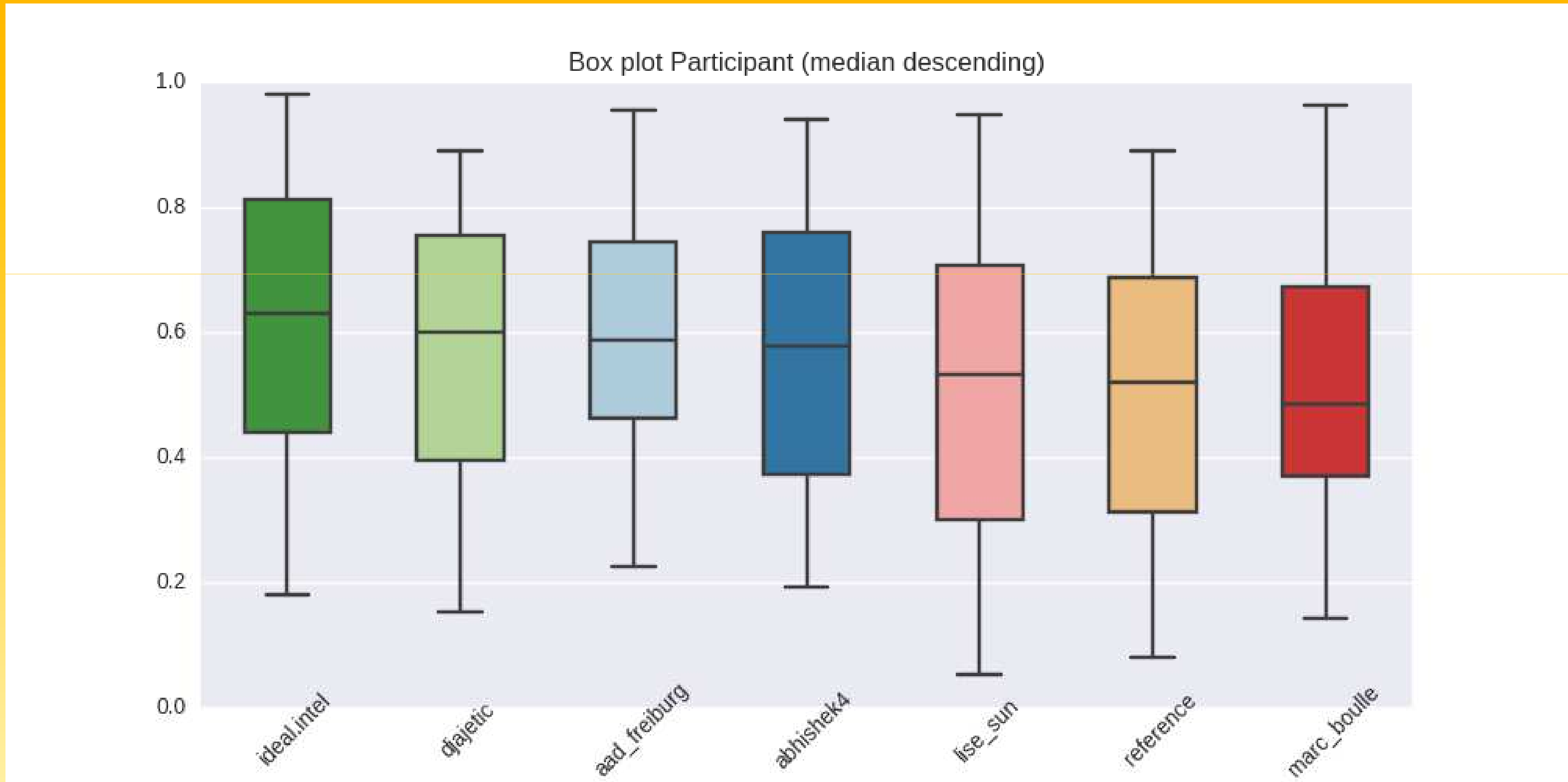
Rnd	AutoML				Final				UP (%)
	Ended	Winners	$< R >$	$< S >$	Ended	Winners	$< R >$	$< S >$	
0	NA	NA	NA	NA	02/14/15	1. ideal 2. abhi 3. aad	1.40 3.60 4.00	0.8159 0.7764 0.7714	NA
1	02/15/15	1. aad 2. jrl44 3. tadej	2.80 3.80 4.20	0.6401 0.6226 0.6456	06/14/15	1. aad 2. ideal 3. amsl	2.20 3.20 4.60	0.7479 0.7324 0.7158	15
2	06/15/15	1. jrl44 2. aad 3. mat	1.80 3.40 4.40	0.4320 0.3529 0.3449	11/14/15	1. ideal 2. djaj 3. aad	2.00 2.20 3.20	0.5180 0.5142 0.4977	35
3	11/15/15	1. djaj 2. NA 3. NA	2.40 NA NA	0.0901 NA NA	02/19/16	1. aad 2. djaj 3. ideal	1.80 2.00 3.80	0.8071 0.7912 0.7547	481
4	02/20/16	1. aad 2. djaj 3. marc	2.20 2.20 2.60	0.3881 0.3841 0.3815	05/1/16	1. aad 2. ideal 3. abhi	1.60 3.60 5.40	0.5238 0.4998 0.4911	31
G P U	NA	NA	NA	NA	05/1/16	1. abhi 2. djaj 3. aad	5.60 6.20 6.20	0.4913 0.4900 0.4884	NA
5	05/1/16	1. aad 2. djaj 3. post	1.60 2.60 4.60	0.5282 0.5379 0.4150	NA	NA	NA	NA	NA

aad=aad_freiburg
abhi=abhishek4
asml=amsl.intel.com
djaj=djajetic
ideal=ideal.intel.analytics
jlr44 = backstreet.bayes
marc=marc.boulle
mat=matthias.vonrohr
post = postech.mlg_exbrain
tadej=tadejs

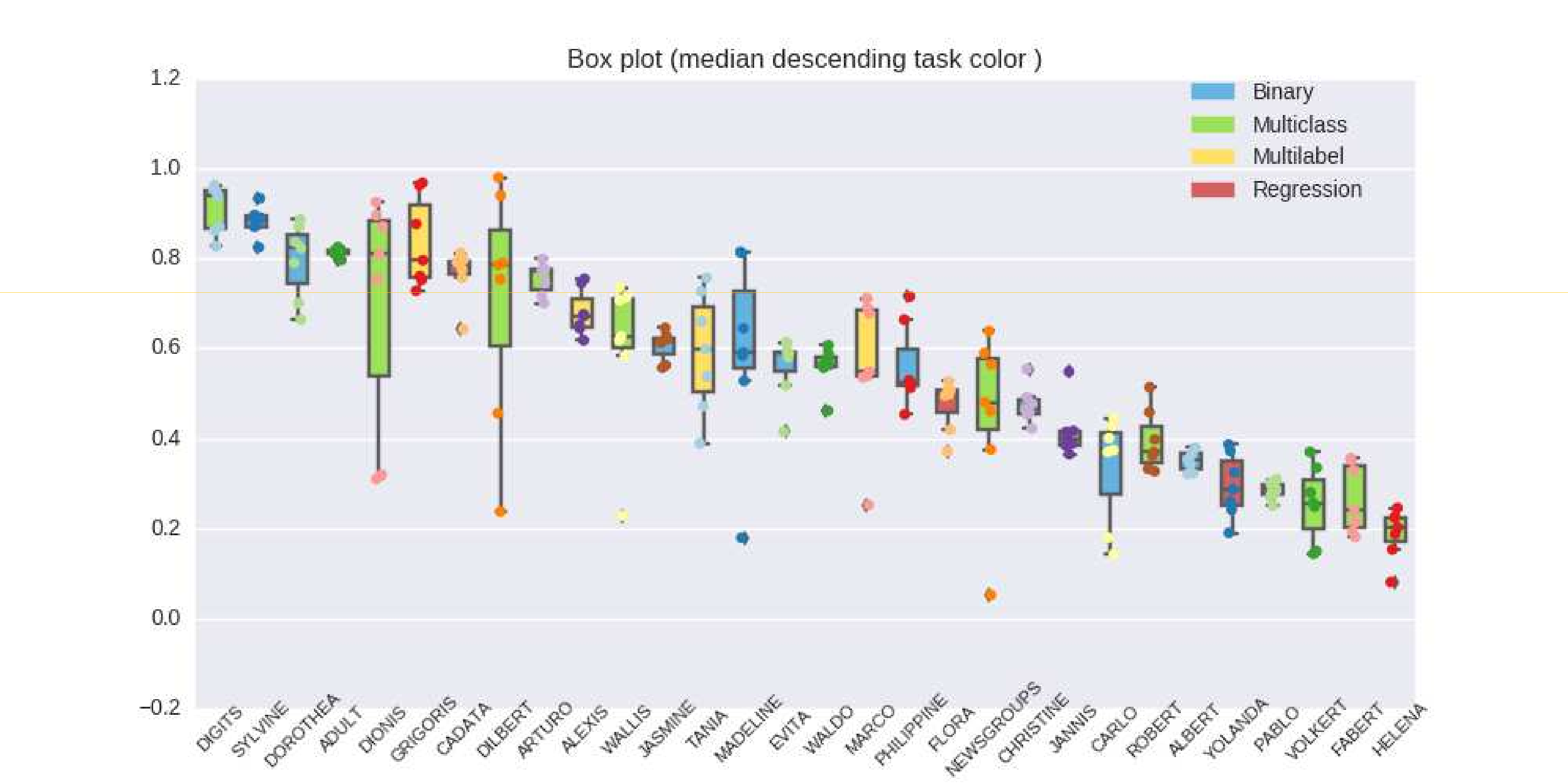
POST-CHALLENGE ANALYSIS

Last submitted code

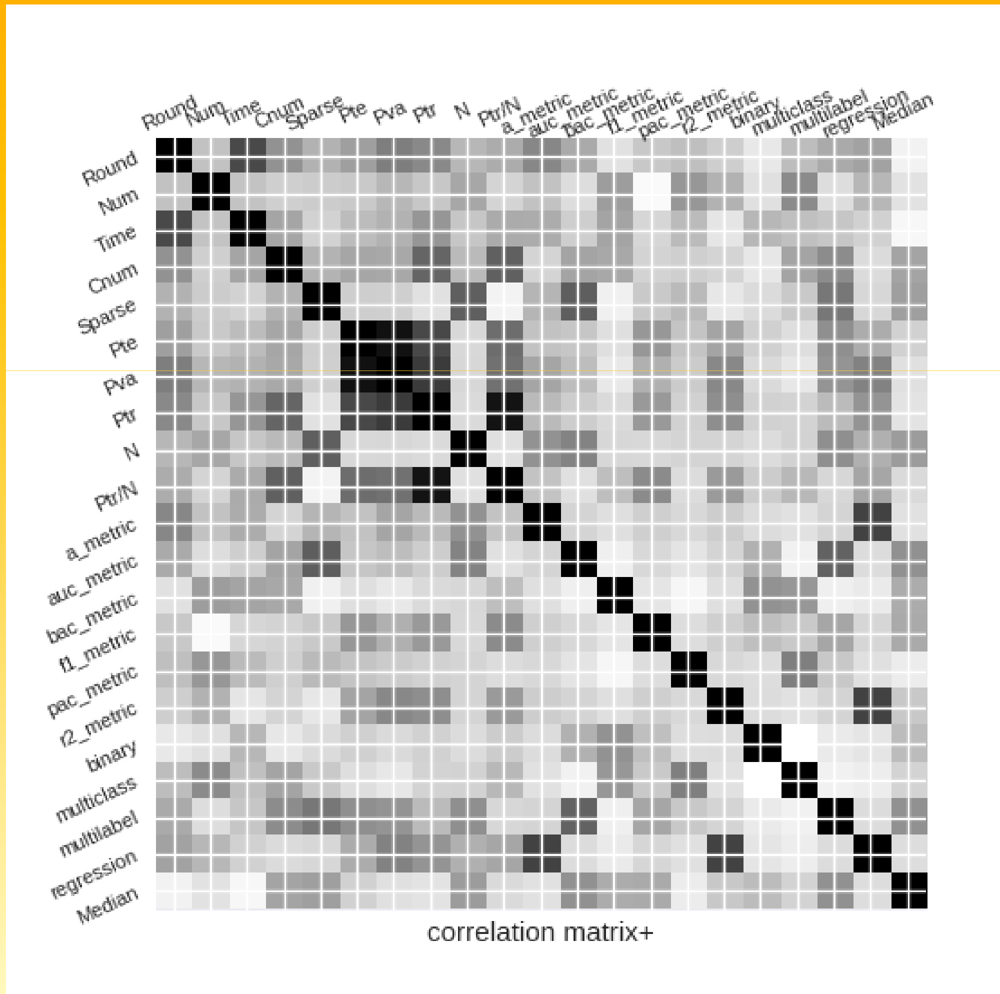
average on all 30 datasets



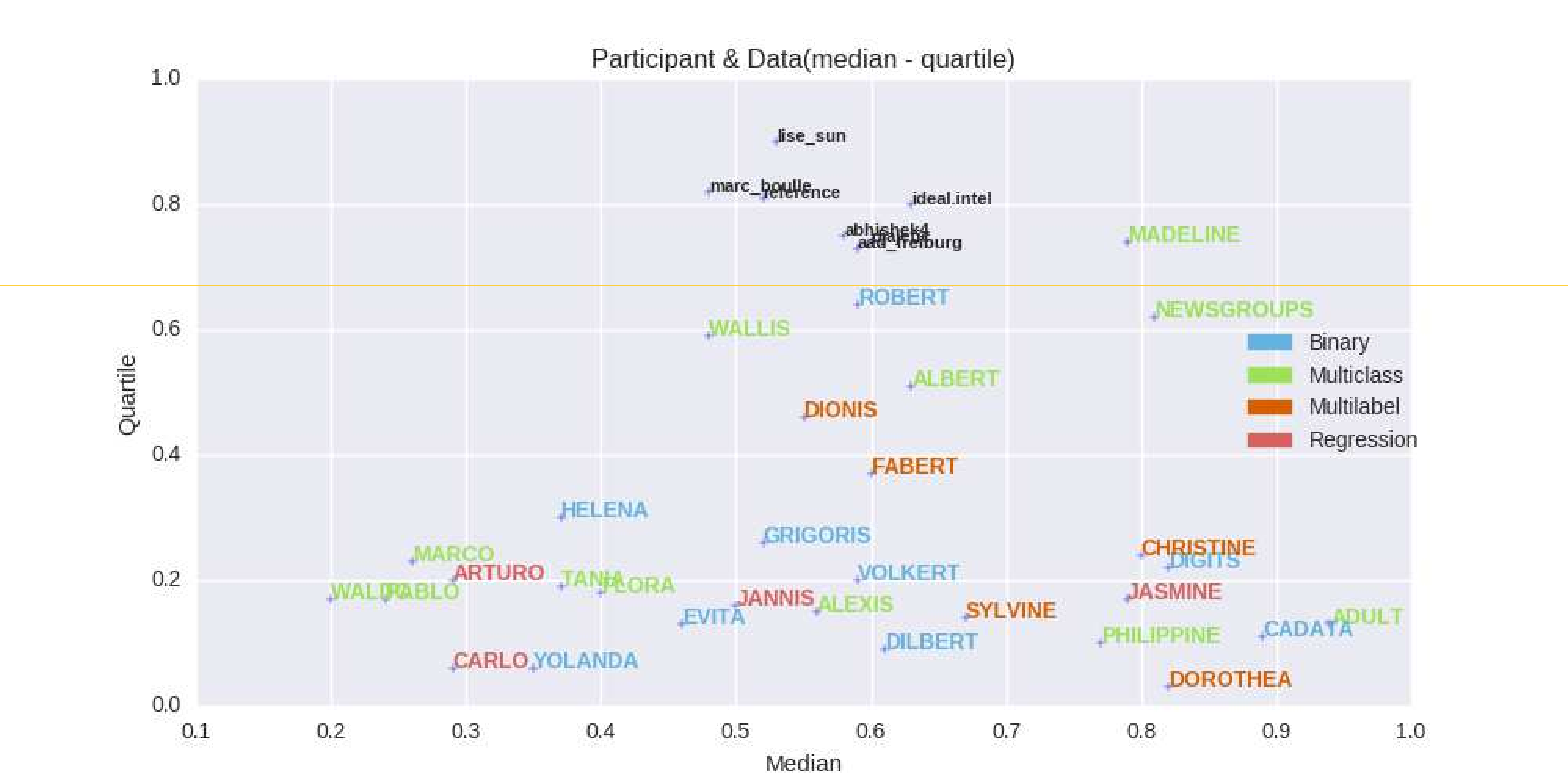
Results per dataset



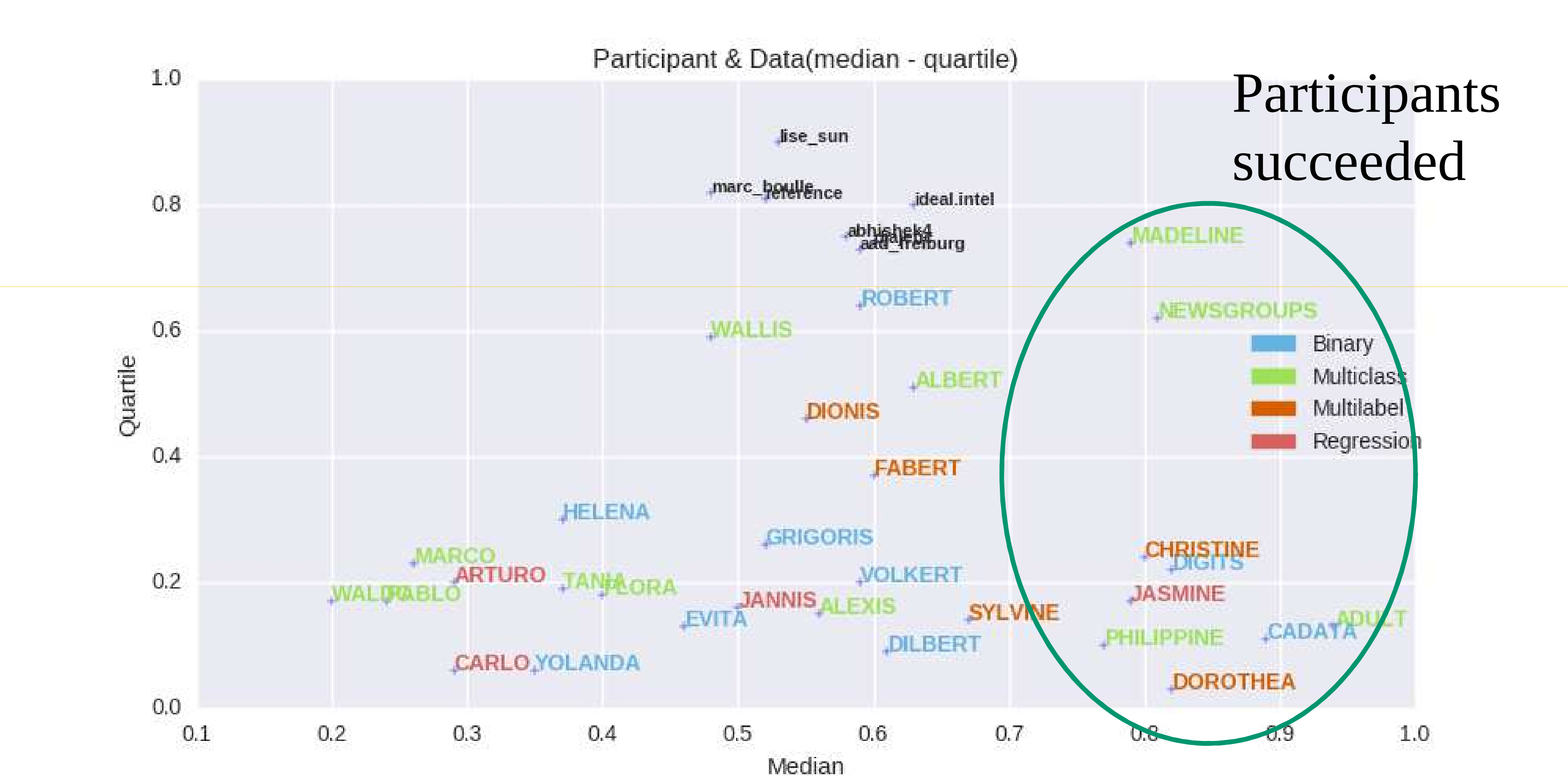
Nothing salient



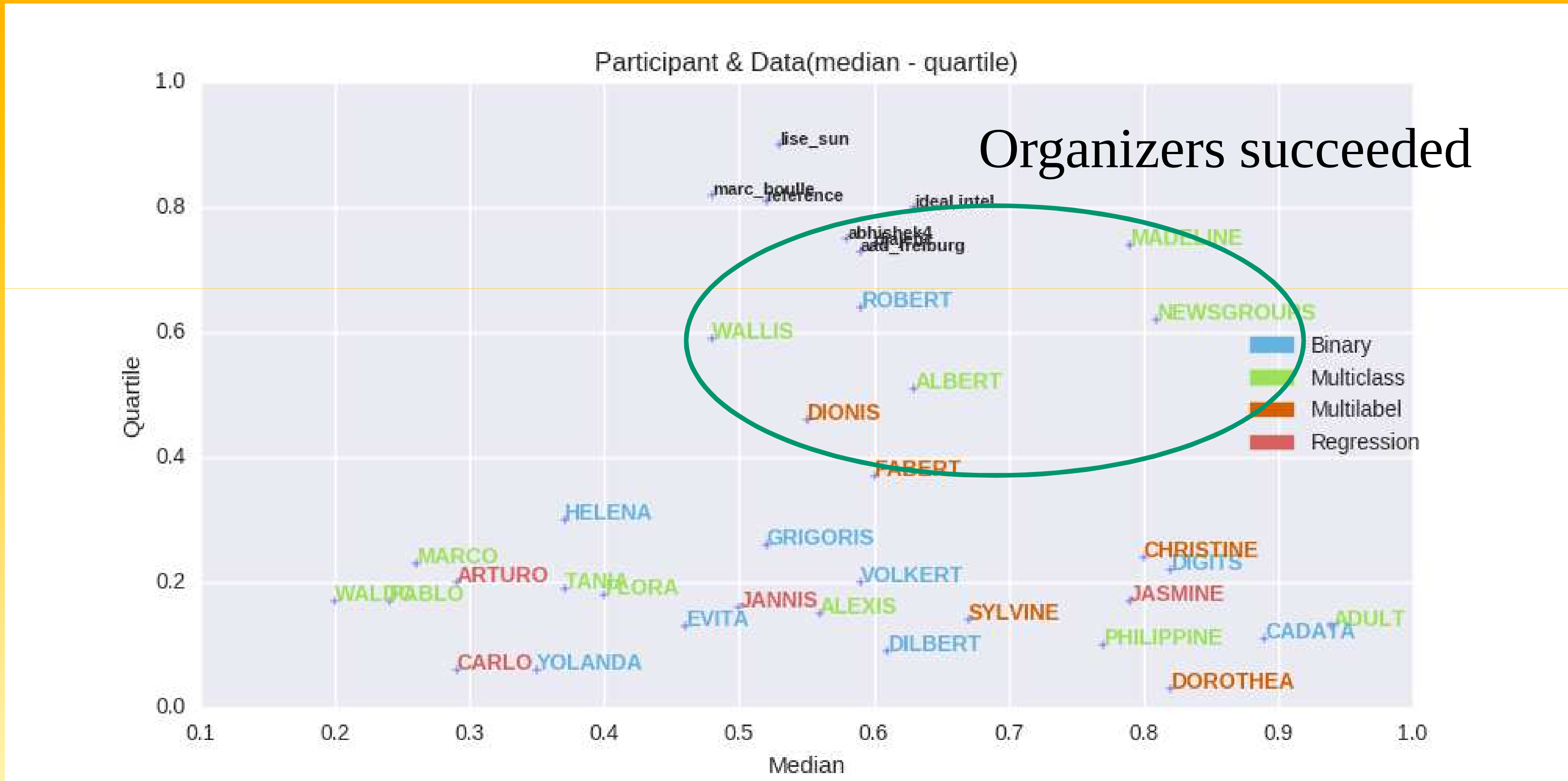
Adversarial problem…



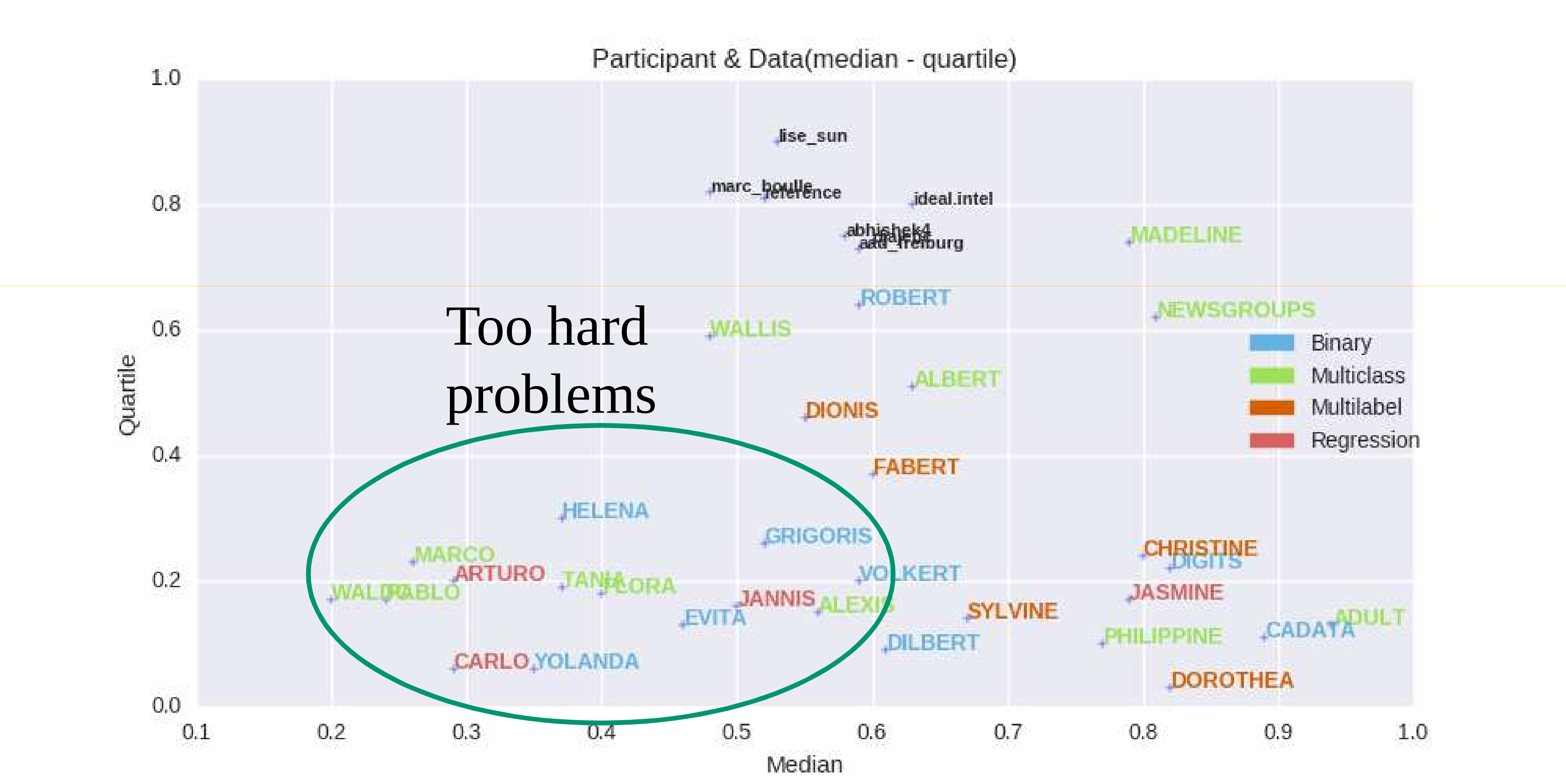
Adversarial problem…



Adversarial problem…

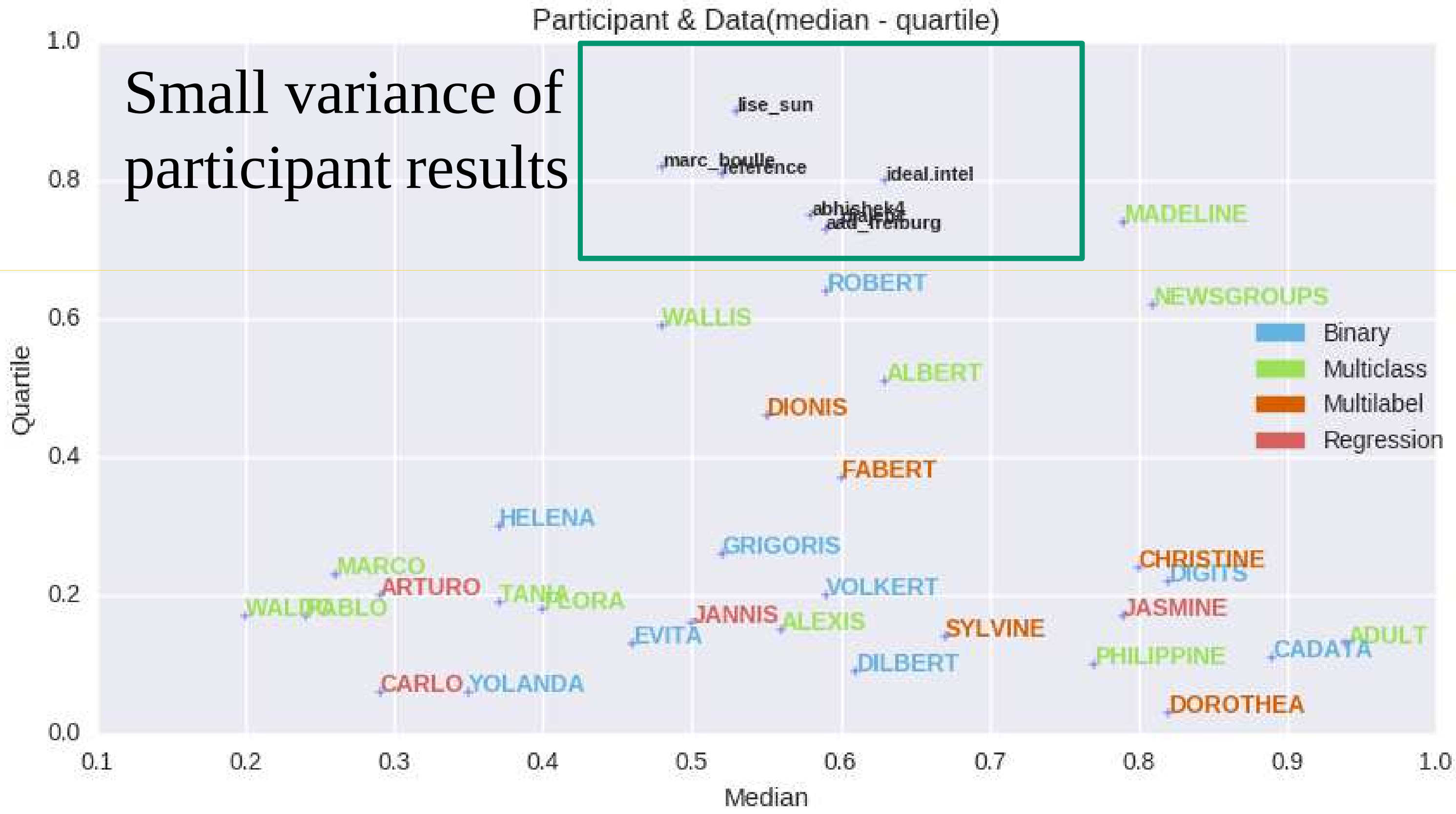


Adversarial problem…

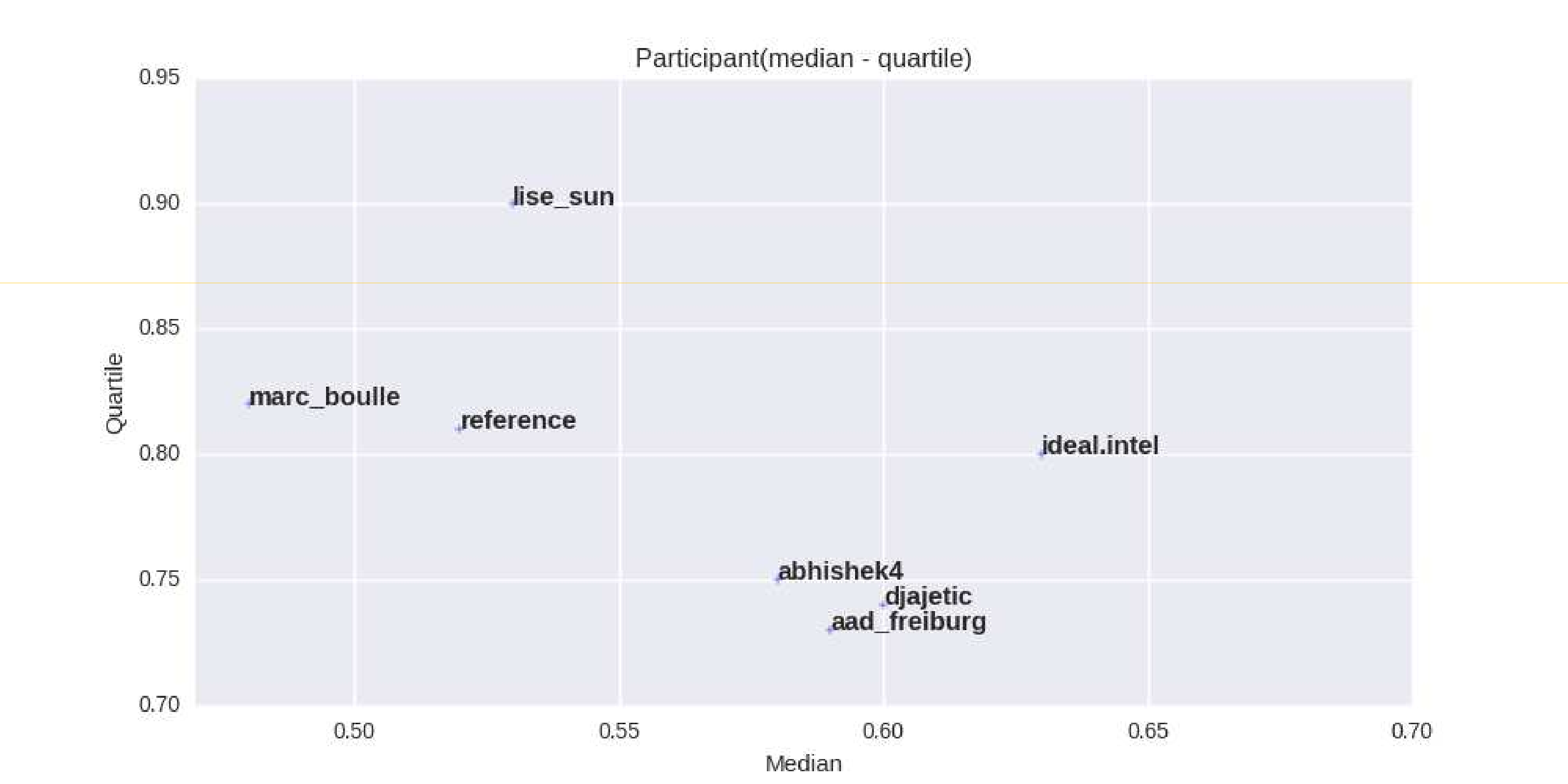


Adversarial problem…

Small variance of participant results



Heterogenous ensembles or single type of classifier?



Conclusion

- We have made great progress, but there is still a “gap”: Tweakathons get about 30% better than AutoML.
- Free Auto-sklearn thanks to aad_freiburg (thanks!)
- The gap is not so big, we are almost there, but we could make the tasks harder (unpreprocessed data, non i.i.d. data, etc.)
- The challenge is hopefully the first in a series: moving towards life long AutoML and adversarial AutoML.

More on http://automl.chalearn.org/

Beat AutoSKlearn game @WCCI16

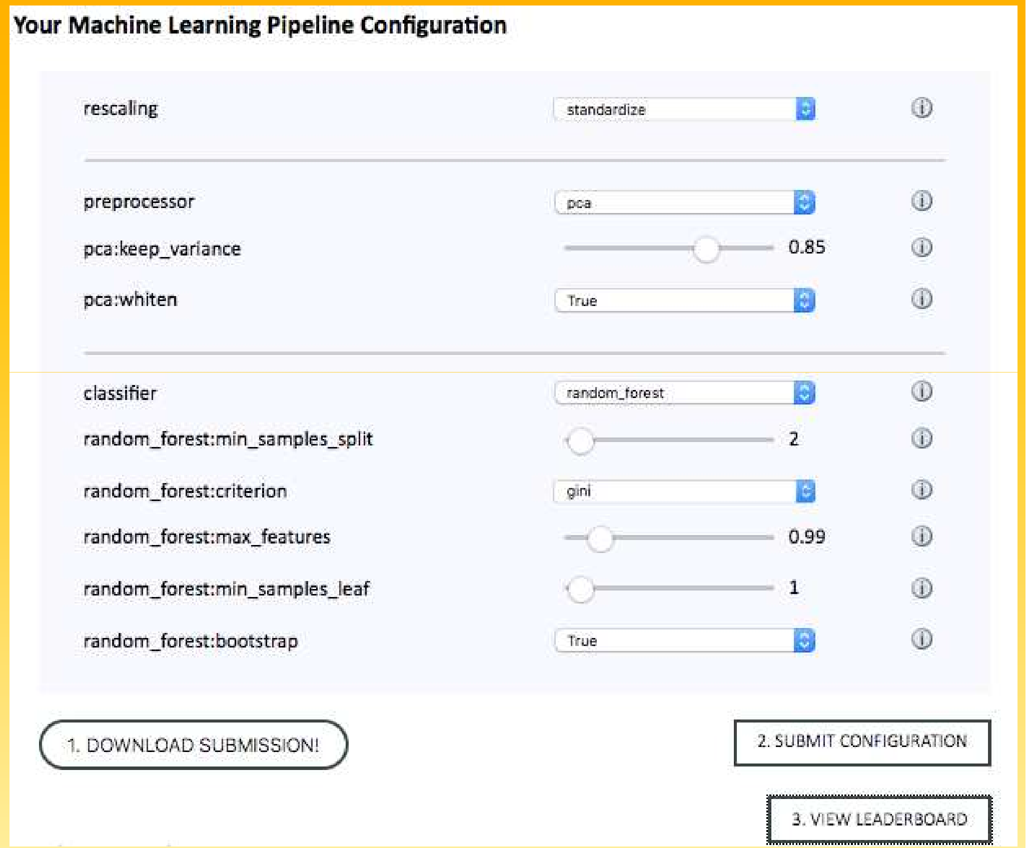
- July 24 - July 26, 2016:
  Fun machine learning game!

- Try to beat the winners of the AutoML challenge by manually tweaking hyper-parameters.

- http://autosklearn.codalab.org

Beat AutoSKlearn game @WCCI16

- A friendly GUI for (manual) configuration generation & submission was provided by the aad_freiburg team and ChaLearn



Was AutoSKlearn beat?

- No! …… Yet

RESULTS

	User	Score
1	AUTOSKLEARN	0.7815 (1)
2	borbudak	0.7544 (2)
3	lisa	0.7507 (3)
4	brunoseznec	0.7372 (4)
5	djajetic	0.6921 (5)
6	Krxsy	0.6643 (6)
7	finlouarn2	0.6502 (7)
8	FlyingFox	0.6350 (8)
9	smohsinali	0.6108 (9)
10	lise_sun	0.5989 (10)
11	aaron	0.4935 (11)
12	hugo.jair	0.4585 (12)
13	Guru	0.3905 (13)
14	pellegrin	0.1957 (14)
15	a.morales	0.1245 (15)
16	aems30	0.0958 (16)

Was AutoSKlearn beat?

- The game was open for a very short period (48hours)
- Half of the participants submitted random-models
- Top ranked participants were very close to the AutoML winner, yet it was not beaten
- It will be interesting to open the game for more time and give it more dissemination so that AutoSKlearn can be really challenged!

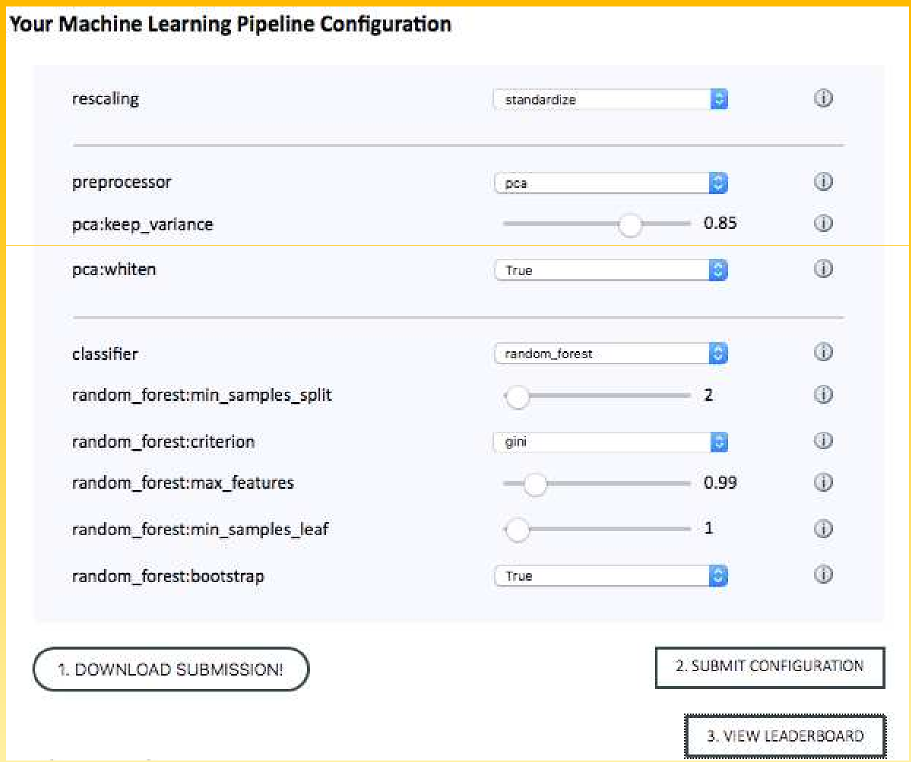
Beat AutoSKlearn game @WCCI16

- Will remain open for a few days
- We will organize a coopetition around it
- Please participate, disseminate!
  - http://autosklearn.codalab.org

Results of the AutoML challenge

- ***Isabelle Guyon, Imad Chaabane, Hugo Jair Escalante, Sergio Escalera, Damir Jajetic, James Robert Lloyd, Nuria Macia, Bisakha Ray, Lukasz Romazco, Michele Sebag, Alexander Statnikov, Sebastien Treger, Evelyne Viegas***
- *automl@**ChaLearn**.org*

Beat AutoSKlearn game @WCCI16




IEEE World Congress on Computational Intelligence
24-29 July 2016, Vancouver, Canada

Beat Auto-Sklearn @ WCCI2016

July 24 - July 26, 2016: Fun machine learning game!
Try to beat the winners of the AutoML challenge by manually tweaking hyper-parameters.

AutoSklearn is a novel solution for autonomous machine learning that recently won the ChaLearn AutoML competition,. AutoML focused in developing automatic methods to solve a variety of supervised learning problems ***without any*** human intervention. Methods for AutoML, including AutoSklearn, aim at taking the human expert out of the loop, are you going to allow this?

I WANT YOU TO BEAT AUTO-SKLEARN

Please visit the challenge website to participate
(it's embarrassingly easy):
http://autosklearn.codalab.org

The challenge will be open to anyone from July 24 to July 26 (results will be announced on July 26 during the contest session of WCCI AutoML Challenge II at Vancouver, Canada).

Contact:
autosklearn@chalearn.org

CHA LEARN ML4AAD